Defense Technical Information Center Compilation Part Notice

ADP014604

TITLE: Comments on Presentation by Paul Cox

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: Proceedings of the Eighth Conference on the Design of Experiments in Army Research Development and Testing

To order the complete compilation report, use: ADA419759

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, etc. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:
ADP014598 thru ADP014630

COMMENTS ON PRESENTATION BY PAUL COX

H. L. Lucas
Institute of Statistics
North Carolina State College
Raleigh, North Carolina

The panelists so far have covered just about everything that I had in mind. I certainly agree in the main with the comments they have made regarding just what particular points on the observed curve or what particular function of the observations may be of interest. Also, I agree with the comments regarding the desirability of fitting a "rational" model, which presumably can be supplied, at least in approximate form, by the engineers. I wish, however to expand on a very important point.

Many of the remarks of panelists about design and analysis have been engendered by the existence of "noise" along the curve for an individual motor and the probable lack of independence of successive observations. I wish to emphasize that there is another, and probably much more important, "noise" component involved. The latter arises from the fact that a group of motors which are constructed and treated alike, insofar as can be managed, will, nevertheless, have inherently somewhat different curves. That is, there is "between-motor" noise as well as "within-motor" (along-the-curve) noise. The existence of between-motor noise must be taken into account for proper experiment design and analysis.

It is instructive to formalize the situation in a way which encompases the two noise components. For the jth experimental unit (here the motor, but in other cases a machine or an animal, etc.) on the ith treatment, we can write the model,

$$y_{ij}(t) = \phi(t; \underline{\theta}_{ij}) + \epsilon_{ij}(t) \tag{1}$$

where

$y_{ij}(t)$ = observed time curve for the unit

$\phi(t;\underline{\theta}_{ij})$ = "true" time curve for the unit

$\underline{\theta}_{ij}$ = vector of parameters for the unit

$\epsilon_{ij}(t)$ = "within-unit" noise,

For the jth unit on the ith treatment, we next write

$$\underline{\theta}_{ij} = \underline{\theta}_i^* + \underline{\delta}_{ij} \tag{2}$$

where

$\underline{\theta}_i^*$ = expected value of $\underline{\theta}_{ij}$ for units on the ith treatment

$\underline{\delta}_{ij}$ = "between-unit" noise.

Substituting (2) into (1) yields the model desired, namely,

$$y_{ij}(t) = \phi\left[t; (\underline{\theta}_i^* + \underline{\delta}_{ij})\right] + \epsilon_{ij}(t), \tag{3}$$

Suppose we compute $\underline{\theta}_{ij}$, an estimate of $\underline{\theta}_{ij}$, for each unit. We see that

$$\hat{\underline{\theta}}_{ij} = \underline{\theta}_{ij} + \underline{\eta}_{ij} \tag{4}$$

where

$\underline{\eta}_{ij} = \underline{\eta}\left[t; \underline{\theta}_{ij}; \epsilon_{ij}(t)\right]$, a vector of errors with which $\underline{\theta}_{ij}$ is estimated; these stem from "within-unit" noise.

We are interested, however, in estimating $\underline{\theta}_i^*$. The relation of $\hat{\underline{\theta}}_{ij}$ to $\underline{\theta}_i^*$ can be seen by substituting (2) into (4) to obtain

$$\begin{aligned} \hat{\underline{\theta}}_{ij} &= \underline{\theta}_i^* + \underline{\delta}_{ij} + \underline{\eta}_{ij} \\ &= \underline{\theta}_i^* + \underline{\delta}_{ij} . \end{aligned} \tag{5}$$

Note that $\underline{\delta}_{ij}^{*} = \underline{\delta}_{ij} + \underline{\eta}_{ij}$ is the total noise or error in $\hat{\underline{\theta}}_{ij}$ and stems from both "between" and "within" noise.

In view of the development just completed, it is certainly reasonable first to estimate $\underline{\theta}_{ij}$ for each individual unit and then as a second step, to analyze the $\hat{\underline{\theta}}_{ij}$ according as the experimental design dictates. Since $\hat{\underline{\theta}}_{ij}$ is a vector, multivariate methods may be desired. Note that the procedure is a "robust" one.

Some papers in which the "robust" approach has been employed are [1], [3], [4], [5], [6].

In view of the remarks of some of the other panelists about choice of points along the time curve and about correlation between successive observations along the curve, the following comments seem in order. In my experience, the contribution of the "between" noise, $\underline{\delta}_{ij}$, to the variance of $\hat{\underline{\theta}}_{ij}$ as an estimate of $\underline{\theta}_i^{*}$ is dominant over the contribution of the "within" noise as summed up in $\underline{\eta}_{ij}$. In fact, in some instances, the "between" noise, $\underline{\delta}_{ij}$, is large relative to the "within" noise, $\epsilon_{ij}(t)$, itself; in this event, the contribution of $\underline{\eta}_{ij}$ is negligible. With $\underline{\delta}_{ij}$ dominant over $\underline{\eta}_{ij}$, it is clear that one need not worry much about the correlation between successive observations on the same unit, that any reasonable method of computing $\hat{\underline{\theta}}_{ij}$ will do, and that one needs use only the minimum number of points along the ij^{th} curve consistent with the complexity of ϕ and the obtaining of moderately efficient estimates of $\underline{\theta}_{ij}$.

This leads next to the design problem, a matter which has been discussed by the other panelists primarily from the standpoint of selecting points along the time curve. In view of my foregoing remarks, I cannot see that the pattern for selection of points along the time curve is the really critical matter, just as long as the pattern is a reasonable one. Instead, the important question is how to select an optimum set of treatment combinations.

To comment further about the design problem, it is again advantageous to be somewhat formal. We note that $\underline{\theta}_i^{*}$ is a function of the levels of the

treatment variables (here, temperature and mixture); i.e.,

$$\underline{\theta}_i^* = \underline{\gamma}(\underline{x}_i;\ \underline{\alpha}) \tag{6}$$

where

γ = a vector of functions of the vectors $\underline{x}_i$ and $\underline{\alpha}$

$\underline{x}_i$ = the vector of levels of the treatment variables characterizing the i^{th} treatment;

$\underline{\alpha}$ = a vector of parameters which depends on basic invariants and on the levels maintained for treatment-type factors not under study (i.e., factors held constant over all i).

Substituting (6) into (5) yields

$$\hat{\underline{\theta}}_{ij} = \underline{\gamma}(\underline{x}_i;\ \underline{\alpha}) + \underline{\delta}_{ij}^* . \tag{7}$$

Now, if the functional forms represented by $\underline{\gamma}$ are known, the problem is to select a minimum optimal set of x-vectors such that all elements of $\underline{\alpha}$ can be estimated and that the estimate, $\hat{\underline{\alpha}}$, is "best" in a suitable sense. In general the optimum design depends on $\underline{\alpha}$, but, since $\underline{\alpha}$ is unknown, one must use previous estimates (or best guesses) about $\underline{\alpha}$ in order to arrive at a good design. Some ideas about this problem are given in [2]. If the forms of the functions, $\underline{\gamma}$, are subject to question, the design must have extra x-vectors so that tests about the assumed $\underline{\gamma}$ and insight about improvements can be obtained. The latter point is also discussed briefly in [2].

I have finished the main things I want to say. There are, however, a couple of other matters that come to mind.

The first has to do essentially with what function of ϕ and hence of $y_{ij}(t)$ is really of concern to the investigator. Although, in some instances, only a particular univariate function of ϕ may ever be of interest, my experience indicates that this is not generally true. I suggest, therefore, that ordinarily it will be best to study ϕ; i.e., to fit the parameters, $\underline{\theta}_i^*$,

or more basically, $\underline{a}$. Given such fits, anything desired can be ascertained.

Finally, in the first analysis Mr. Cox outlined, he failed to distinguish "between" and "within" noise. The variance sources for his analysis were

Treatment

Time

Time by treatment

Residual.

They should have been

Treatment

Motor within treatment (Error for treatment; corresponds to δ^*_{ij}

Time

Time by treatment

Time by motor within treatment (Error for time and time by treatment; corresponds to ϵ_{ij}).

In closing, I should note that Mr. Cox, in all but his first analysis, adopted the "robust" approach. I stress the approach, however, because it is important, and because judging from his first analysis, Mr. Cox appeared not to be very clear on the implications of the existence of both "between" and "within" noise.

REFERENCES

[1] Box, G.E.P. 1950. Problems in the Analysis of Growth and Wear Curves. Biometrics 6, No. 4, 362-389.

[2] Box, G.E.P. and H.L. Lucas. 1959. Design of experiments in non-linear situations. Biometrika 46: 77-90.

[3] Elston, R.C. and J.E. Grizzle. 1962. Estimation of Time-Response Curves and Their Confidence Bands. Biometrics 18, 2, 148-159.

[4] Fisher, R.A. 1920. Studies in Crop Variation. I. An Examination of the Yield of Dressed Grain from Broadbalk. Jour. Agric. Sci. 11: 107-135.

[5] Fisher, R.A. 1924. The Influence of Rainfall on the Yield of Wheat at Rothamsted. Phil. Trans. Roy. Soc. B, 213. 89-

[6] Vanasse, N.A., I. D. Jones and H. L. Lucas. 1951. Specific gravity-dry matter relationship in potatoes. Am Potato Jour. 28: 781-791.